PREPRINT UCRL- 82501

Lawrence Livermore Laboratory

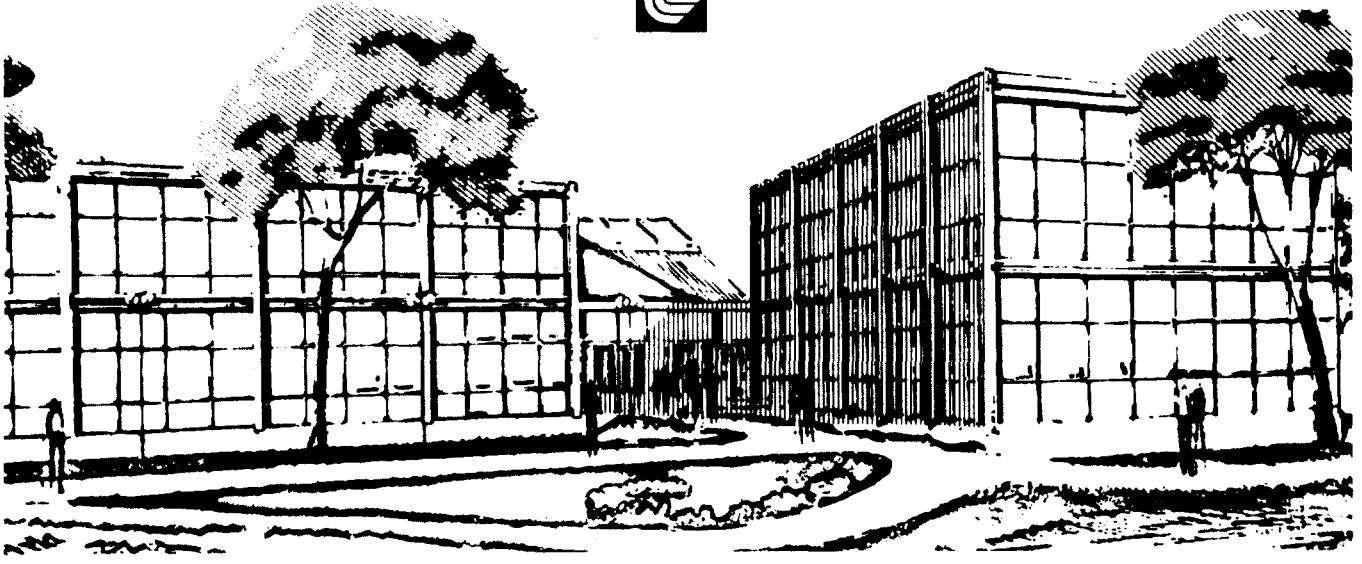
MAHALANOBIS DISTANCE AND VARIABLE SELECTION TO OPTIMIZE DOSE RESPONSE

Dan H. Moore II, Deborah E. Bennett, Andrew J. Wyrobek, and Debbie Kranzler

May 29, 1979

This paper was prepared for submission to

Proceedings of the 1979 annual Meeting of the American Statistical Association and the North American Regions of the Biometric Society, Aug 13-16, Washington,DC



Mahalanobis Distance and Variable Selection to Optimize Dose Response*

by

Dan H. Moore II, Deborah E. Bennett, Andrew J. Wyrobek and Debbie Kranzler, Biomedical Sciences Division, Lawrence Livermore Laboratory, Livermore, California 94550

SUMMARY

A battery of statistical techniques are combined to improve detection of low-level dose response. First, Mahalanobis distances are used to classify objects as normal or abnormal. Then the proportion classified abnormal is regressed on dose. Finally, a subset of regressor variables is selected which maximizes the slope of the dose response line. Use of the techniques is illustrated by application to mouse sperm damaged by low doses of x-rays.

This work was supported by the United States Department of Energy, contract number W-7405-ENG-48 and the Environmental Protection Agency Pass Through Agreement EPA-IAG-D5-E681-AN.

Key words: Dose response, Mahalanobis distance, variable selection.

INTRODUCTION

This study is concerned with developing a statistical methodology which will help to improve the detection of low-level effects of harmful agents. The problem is divided into three parts. First, it is necessary to classify objects exposed to an agent as "normal" or "abnormal" based on a vector of measurements. Initially the objects are unclassified, but for each there is available an independent, discrete measurement (the dose of the agent) which is assumed to be related to the likelihood that the object is abnormal. Next, the relationship between the dose of the agent and the degree of abnormality (the response) is to be quantified. Finally, a subset of variables from the measurement vector is to be selected which optimizes the dose-response relationship. This methodology can be used in measuring the effects of possibly hazardous environmental agents such as air pollution, exposure to chemicals or radiation, or in testing a new drug for possibly harmful side effects.

Our method uses a variety of statistical techniques which are not new but their combination is and has proven useful in a recent practical application. We begin by describing this application.

DESCRIPTION OF THE DATA

Chemical mutagens and x-irradiation affect the morphology

of sperm heads in a way that can be distinguished under a microscope (Wyrobek and Bruce 1978). Normally a biologist studies each sperm under a microscope and makes a subjective judgement as to whether the sperm is "normal" or "abnormal". The percent abnormal is then plotted against dose and used to find an estimate for a "doubling dose", that dose which leads to twice the background (0 dose) abnormal percentage. In our experiment groups of 3 mice received acute, testicular doses of 0, 30, 60, 90 or 120 rads of x-irradiation. For each mouse 50 sperm were chosen at random, photographed and enlarged. Eleven measurements were made on each of the 750 sperm head silhouettes (Figure 1).

Initially the sperm used in this study were not classified by a biologist since our goal was to try to develop a system which is more sensitive than the subjective one currently used. Thus, our first problem was to find a way of relating the measurements to the dose of x-rays.

ESTABLISHING A DOSE-RESPONSE RELATIONSHIP

A useful measure of the difference between a p-variate observation vector $x = (x_1, \ldots, x_p)$ and a group mean vector $\bar{x} = (\bar{x}_1, \ldots, x_p)$ is the Mahalanobis distance (M-distance) defined by

$$M(x) = (x - \bar{x})' S^{-1} (x - \bar{x}) ,$$

where S is the group sample covariance matrix. This measure can

be thought of as a distance in a p-dimensional space which takes into account the scales of the measurements as well as correlations between pairs.

We pooled the 150 observations from 3 mice with 0 rad exposure to form a control group. Each observation can now be expressed as an M-distance from the control group (0 rad) mean. We expect that, on the average, observations from mice receiving high doses of radiation will have greater M-distances than observations from mice receiving low doses. We can also treat the M-distance as a dependent variable and regress it on dose. Figure 2 shows the result when the mean M-distances for the 15 mice are regressed on the 5 dose levels in our experiment. The regression line in the figure has intercept 10.87 and slope 0.14.

A point estimate for the doubling dose is the intercept (A) divided by the slope (B). In this case the estimated doubling-dose is 76 rads. This is roughly equivalent to the 70 rad doubling dose established by the conventional method based on visual scoring of 500 sperm per mouse.

Replicate measurements at each dose can be used to measure the goodness-of-fit of the regression line to this data. This is accomplished through the F-statistic

$$F = \frac{\sum_{i=1}^{k} n_i (\bar{Y}_i - \hat{Y}_i)^2 / (k-2)}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / (N-k)}$$

where Y_{ij} is the response (mean M-distance) of the jth mouse at the ith dose level, $\bar{Y}_i$ the mean response for all 3 mice at the ith dose level, $\hat{Y}_i$ the linear regression predicted response, $n_i = 3$

the number of mice at each dose level, k=5 the number of dose levels and N=15 the total number of mice.

In our case F = 0.77 with 3 and 10 degrees of freedom leading to P = 0.54 for the appropriateness of the linear model. (Subsequently this P value will be referred to as P-linearity.) Although this test indicates that the linear model provides an adequate fit to the data, the estimated doubling dose is no lower than that estimated using current methods.

There is evidence that mean M-distances are not very robust against large "outlier" values from measurements on single sperm. This is shown in the figure where one of the mice at 90 rads has a very large mean M-distance, which was caused by a single outlier among the 50 sperm measurements which contributed to the mean. In addition, biologists are unfamiliar with M-distances and would prefer to see results expressed as percent abnormal sperm. Thus, we seek a method for using M-distances to classify sperm as normal or abnormal. This should also reduce the sensitivity of the measure to large outliers.

CLASSIFICATION OF INDIVIDUAL SPERM

The M-distances for sperm in the control group are identical to the squared radii defined by Gnanadesikan (p.172). Therefore, if we assume that the vector of measurements x has a multinormal distribution, the M-distances will have approximately a chi-squared distribution with degrees of freedom equal to the number of variables. Unfortunately, our data contain too many

"outliers" with very large M-distances for the assumption of multinormality to hold so that we cannot use chi-square critical values. However, we can find a value which will arbitrarily classify a proportion of the control group sperm as abnormal. If the value 28 is chosen 7 of the 150 M-distances for control sperm (4.67%) will be classified as abnormal. This compares with the upper 5% chi-square value (with 11 d.f.) of 19.675. When this value is used to classify the sperm, and the resulting mean percent abnormal for each mouse is regressed on dose, we obtain the result shown in Figure 3. We see a dramatic reduction of the doubling dose to 34 rads and an increase in P-linearity to 0.98. In this case detection of low-level effects has been improved by using M-distances for classification rather than as a quantitative measure of abnormality.

REDUCING THE NUMBER OF MEASUREMENTS

Now we wish to determine whether any significant loss in detection ability occurs when fewer variables are used. Standard techniques for comparisons among subsets of variables cannot be applied here since the dependent variable (percent classified abnormal) is not fixed. It varies depending on the critical value selected for M-distances which, in turn, depends on the number of variables included. With the help of Jane Gentleman's subroutine, ALLNR, it is easy (although time consuming) to compute regressions using all possible combinations of subsets of vari-

ables. With eleven variables this requires computing

$$\binom{11}{11} + \binom{11}{10} + \ldots + \binom{11}{1} = 2^{11} - 1 = 2047$$

regressions. This is easily accomplished on a computer. For each subset of variables a critical value was selected for M-distances which arbitrarily classified 5 % of the zero-rad sperm as abnormal. This facilitates comparisons between the resulting regressions. In comparing regressions we restricted attention to subsets for which P-linearity was 0.8 or higher (guaranteeing that linear regression gives a good fit to the data) and for which the estimated doubling dose is within the upper 95% confidence limit of that found by using all eleven variables. This limit can be found as follows: A point estimate for doubling-dose is A/B which has variance approximately equal to

$$\text{var}(A/B) = \frac{\text{var } A}{B^2} + \frac{A^2 \text{ var } B}{B^4} - \frac{2A \text{ cov}(A,B)}{B^3}$$

(Kendall & Stuart,Vol. 1,p.232). Under the assumption that A/B has an approximate normal distribution, an approximation for the upper 5% confidence limit is given by A/B + t * s(A/B), where t is the upper 5% point of a t-distribution with (n-2) d.f. and s(A/B) is the estimated standard deviation of the doubling dose.

In our case n=15 means so that d.f.=13. Thus, the upper 95% confidence limit for our estimated doubling dose is 55.87. Figure 4 shows all combinations of variables which satisfy the twin criteria P-linearity > 0.8 and doubling-dose < 55.

Most of the points in Figure 4 represent combinations of variables which statistically are equivalent. However, we can reduce this set by noting that the ideal regression has P-linearity = 1.0 and doubling dose = 0. We prefer points near these values over those farther away. A convex hull of admissible points may be constructed by connecting those sharing equal numbers of variables and nearest to the (1.0,0) corner of the figure. Admissible points are defined as those for which there are no other points (for the same number of variables) below and to the right of them. A point is inadmissible if another point (with the same number of variables) has better P-linearity and lower doubling dose. Figure 5 shows convex hulls of admissible points.

COMPARISONS AMONG REGRESSIONS

Admissible points may also be compared. First, we convert the doubling dose to a p-value based on a test of how significantly better it is than the point using all 11 variables. A rough test of this is given by

$$t = \frac{\text{Doubling dose (subset S)} - \text{Doubling dose (all 11 variables)}}{\text{Standard error (Doubling dose subset S)}}$$

which will have an approximate t-distribution with (n-2) d.f. A simple method for combining the two measures of merit is the optimality coefficient defined by

$$\text{Optimality Coefficient} = \sqrt{P_D^2 + (1 - P_L)^2} .$$

where P_D is the P-value for doubling dose test and P_L is P-linearity. An optimality coefficient can be determined for each point, and for each subset of k variables there will be a minimal optimality coefficient.

A plot of these minima against the number of variables is shown in Figure 6. The figure shows that adding variables improves performance, as measured by the optimality coefficient, up to five variables. If more than five variables are used performance deteriorates. Thus, we are able to find an "optimal" subset of variables. This optimal subset is at least no worse than all other subsets and may be better (with probability greater than zero).

COMPARISON WITH CURRENT CLASSIFICATION METHOD

It is interesting to compare classification based on M-distances with subjective classification by experienced biologists. The biologists were asked to classify a subset of the 750 pictures of sperm heads without knowing the M-distances or computer classification. A sample of 100 sperm were selected by the statisticians and classified by the biologists with the results shown in Table I. In this table M-distances are based on the five best variables, as determined by minimizing the optimality coefficient. The table shows that all samples classified as abnormal have M-distances greater than 12. The majority of the

normal samples (67 out of 84 or 80%) have M-distances less than 12. The interesting question is whether the 17 sperm classified as normal by biologists but with M-distances greater than 12 are normal or abnormal. Visually they cannot be distinguished from other normal specimens, but quantitatively they differ from the bulk of the normals.

DISCUSSION

We believe that the idea of using M-distances from a control group as a measure of response is a new and useful one. It can be applied to many situations, particularly those where the exact nature of the response cannot be predicted or described prior to the experiment. In general, M-distances will not be normally distributed so that it will be necessary to transform them prior to quantifying the dose-response relationship. We found it helpful to use M-distances to classify objects as normal or abnormal and to regress the percent classified abnormal on dose. This may not be appropriate in other situations but it does succeed in removing the effects of extreme M-distances on the regression. In our case it allowed us to express the results in terms familiar to biologists.

A second new idea is the use of two criteria to select a subset of variables which maximize the dose-response relationship. The criterion of goodness-of-fit seems a natural one. It is also natural to seek sets of variables which minimize doubling

dose. Statisticians may prefer to select subsets based on the criterion of slope of the regression line, whose standard error can be determined explicitly rather than by approximation. We chose to use doubling dose for two reasons. First, doubling dose is familiar to biologists and results expressed in its terms can be compared with current capabilities. Second, the standard error for doubling-dose, which is used in determining the P-value used in comparisons, includes uncertainties due to both slope and intercept; thus it may provide a more reliable guide than using slope alone as a criterion.

There are many ways the twin criteria of P-linearity and P-doubling dose could have been combined; we chose one that measures the two-dimensional Euclidean distance from the optimal point. Various weighting schemes could be applied, depending on whether linearity is more important or less important than reduction in doubling dose. We only wish to suggest a useful method for selecting a subset of variables when standard methods cannot be applied due to varying regressions.

Our results suggest that careful measurement of sperm head dimensions combined with application of the statistical methods described here can lead to increased detection of low-level effects. This is due to the increased sensitivity of a quantitative measurement system over a subjective, visual one.

ACKNOWLEDGEMENTS

We especially thank F. Ochoa, L. Gordon and G. Watchmaker for their technical assistance in this study.

REFERENCES

1. Gnanadesikan, R. Methods for Statistical Data Analysis of Multivariate Observations, John Wiley & Sons, N.Y. 1977.

2. Gentleman, Jane F. Generation of all ${}_{N}C_{R}$ combinations by simulating nested Fortran DO loops, Algorithm AS 88, Applied Statistics 24:374-6, 1975.

3. Kendall, Maurice G. and Alan Stuart. The Advanced Theory of Statistics, Volume 1, Third Ed., Hafner Publishing Co. N.Y. 1969.

4. Wyrobek,Andrew J. and W. Robert Bruce. The induction of sperm-shape abnormalities in mice and humans, In: Chemical Mutagens, Volume 5 (A. Hollander and F. J. de Serres, Ed.), Plenum Publishing Corp.,pp. 257-285, 1978.

Table 1. Distribution of M-distances in a Sample of 100 Sperm

M-Distance*	Biologists' Classification Normal	Abnormal
<12	67	0
12-14	6	2
14-16	6	1
16-18	1	2
18-20	0	1
20-22	1	1
22-24	2	0
24-26	0	0
26-28	0	1
28-30	0	1
>30	1	7
Totals	84	16

* M-distance based on five "best" variables (L1, L2, L4, W1, Area)

LEGENDS TO FIGURES

Figure 1. Enlarged drawing of mouse sperm head and 11 measurements.

Figure 2. Regression of mean Mahalanobis distance on dose in rads. Mahalanobis distance is based on the 11 dimensions of Figure 1 and and measured from the mean for a control group of 150 sperm from three mice exposed to zero rads. Means are based on 50 sperm from each of three mice at each dose level.

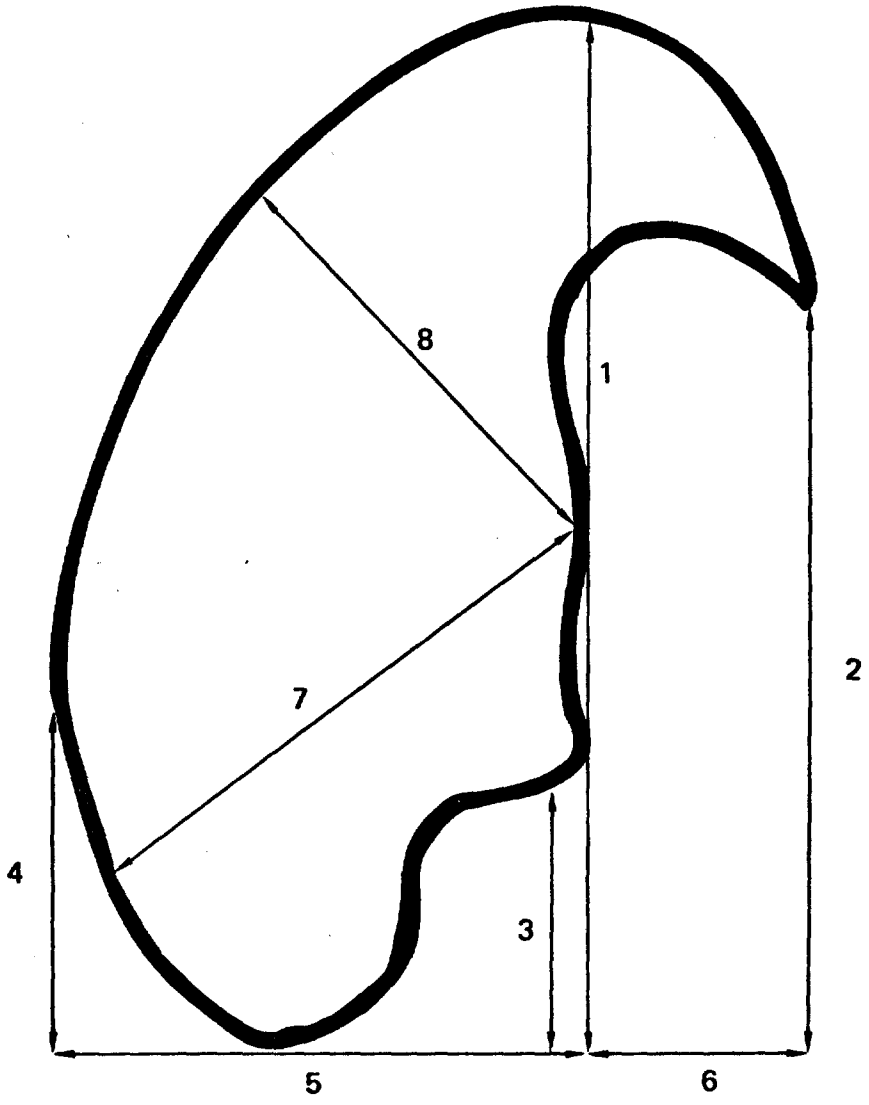
Figure 3. Regression of percent abnormal sperm on dose in rads. Sperm were classified as abnormal if their Mahalanobis distance is greater than 28. The doubling dose estimated from this regression line is 34.

Figure 4. Plots of doubling dose vs. P-linearity for regressions satisfying the criteria: P-linearity > 0.8 and estimated doubling dose < 55. Each point represents a subset of the 11 variables shown in Figure 1. All points are statistically equivalent but we prefer those with low doubling dose and high P-linearity.

Figure 5. Convex hulls of admissible points. These points are a subset of those in Figure 4 and include only those for which there is no point, with the same number of variables, which is both below and to the right of it.

Figure 6. Optimum subsets of variables as determined by the op-

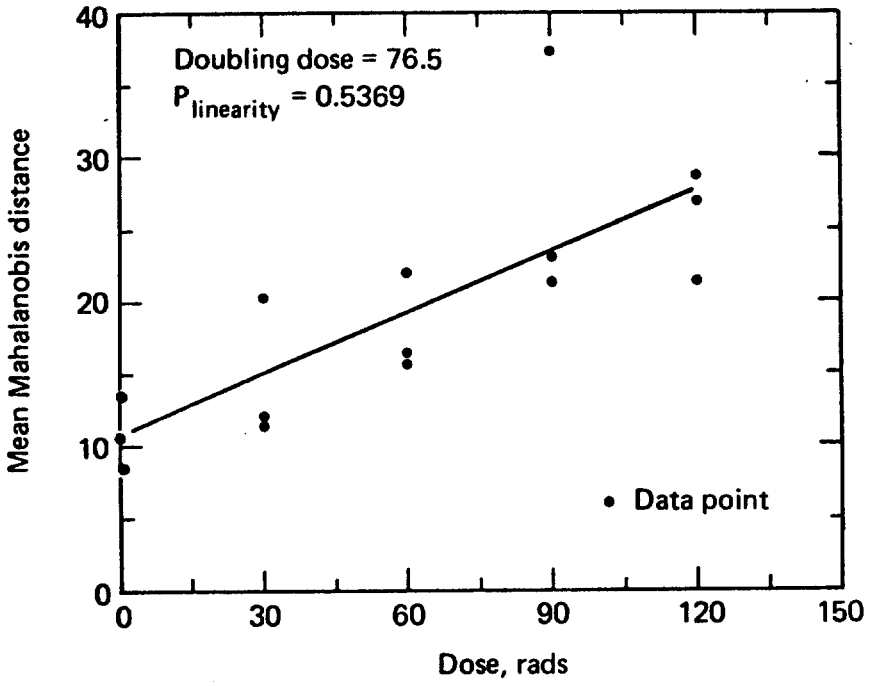
timality coefficient (OC). The OC is equal to the minimum Euclidean distance between points on the convex hulls of Figure 5 and the "ideal" point (1.0,0). Doubling dose has been converted to a P-value prior to measuring the OC. The five variables (1,2,4,5,9) have the lowest OC and represent our best estimate of the most sensitive subset of variables.



1. L_1 – length along axis
2. L_2 – length to tip of hook
3. L_3 – length to tail attachment site
4. L_4 – length to point of maximum width
5. W_1 – maximum width
6. W_2 – width to tip of hook
7. D_1 – lower diagonal
8. D_2 – upper diagonal
9. Area
10. Perimeter
11. Shape = $(\text{Perimeter})^2/(4\pi * \text{Area})$

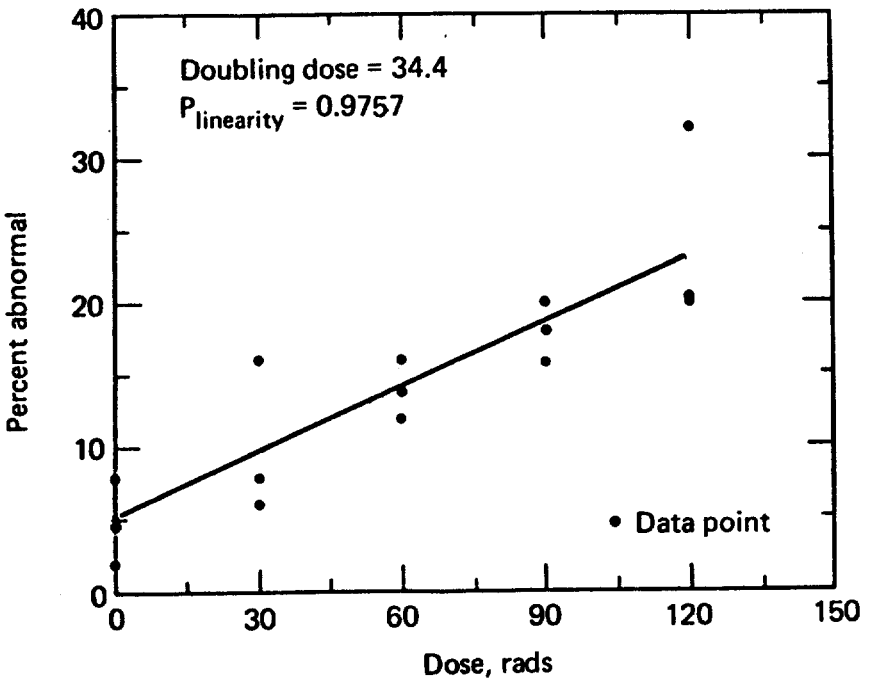
Dan Moore, et al.

Fig. 1



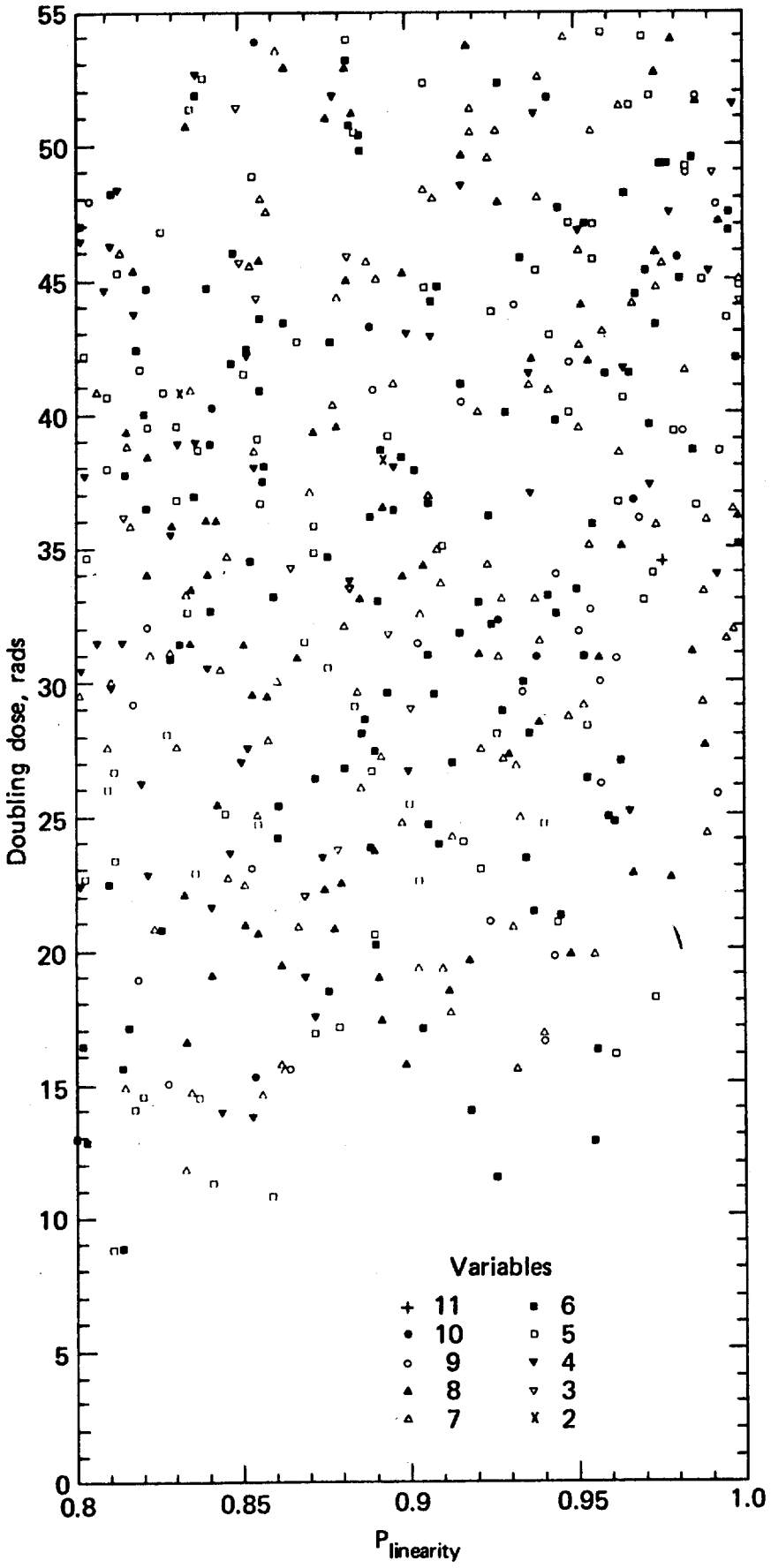
Dan Moore, et al.

Fig. 2



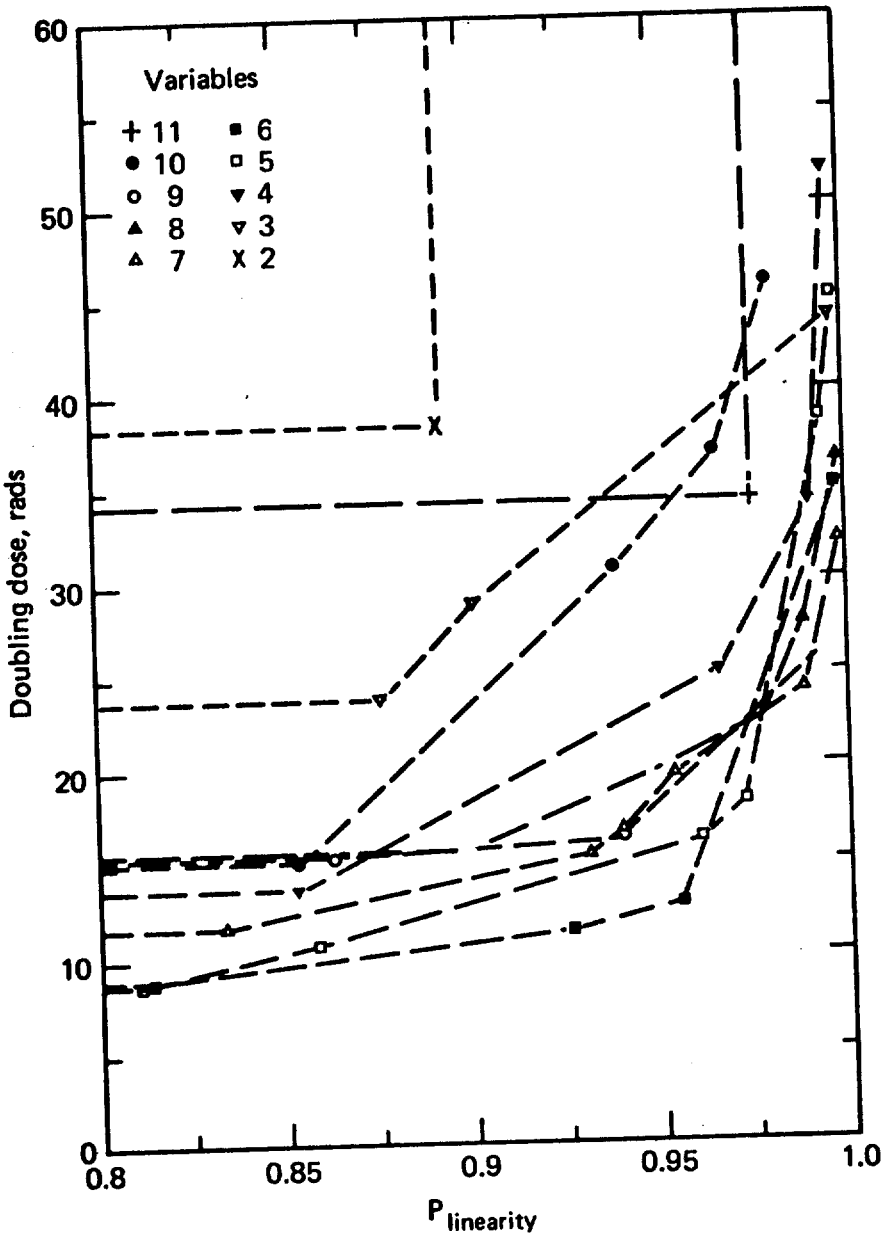
Dan Moore, et al.

Fig. 3



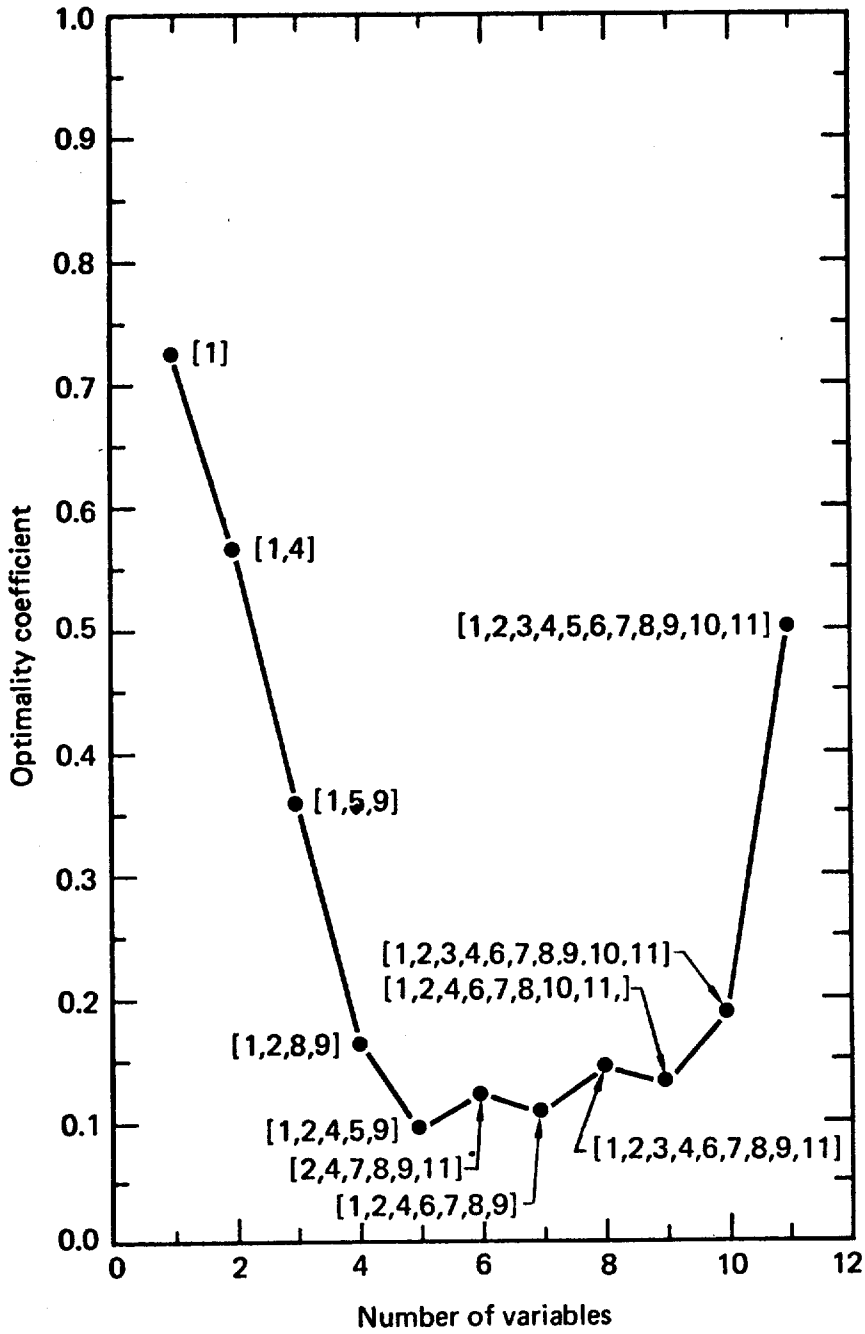
Dan Moore, et al.

Fig. 4



Dan Moore, et al.

Fig. 5



Dan Moore, et al.

Fig. 6